Data Visualization and Analysis (CSE-6242)

egriffin33@gatech.edu (Eric H. Griffin)

June 2, 2018

Activity2

Gun violence Incidents in continential US (2013 ~ 2018)

The goal in this activity is to explore a geolocated dataset), where items are associated with a position on a planet's surface (typically specified using latitude and longitude).

This version is my original rmd that acts as my notebook to show my data analysis (see ac2.pdf for my submitted version which meets grading specifications

The first dataset I chose was listed under the SocialSciences section the following public-datasets, https://github.com/awesomedata/awesome-public-datasets

Additonally a second dataset was used to apply US 2013 Census poulation data into the analysis, this data was created from http://www.enchantedlearning.com/usa/states/population.shtml

I have performed exploratory analysis of the data to compare how states normally attributed to having high concentrations of gun violence that are directly associated to large metropolitan areas is altered when the events are adjusted (per capita) for the total population of the state.

several maps are used to visualize and illustrate total gun death, total injuries and per capita gun deaths.

load map and plotting libraries

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(maps)
library(mapdata)
```

```
# Gun violence dataset for years (2013 ~ 2018)
# dataset obtained from https://github.com/jamesqo/gun-violence-data
# A comprehensive, accessible database that contains records of over
#  260k US gun violence incidents from January 2013 to March 2018.
```

```
gv_data       <- read.csv("stage3.csv", header = TRUE)

#clean dataset, align map state names to gun violence by making all states lowercase
gv_data$state <- tolower(gv_data$state)
```

```
#load plot of United States
states_map <- map_data("state")

# group gun violence data by state summarizing  number killed and number injured
gv_states2    <- gv_data  %>% group_by(state ) %>%
                          summarise(Killed = sum(n_killed), Injured = sum(n_injured))

#create map of actual people killed in gun violence grouped  by state
gv_map        <- merge(states_map, gv_states2, by.x = "region", by.y = "state")
```

```
print(gv_states2)
```

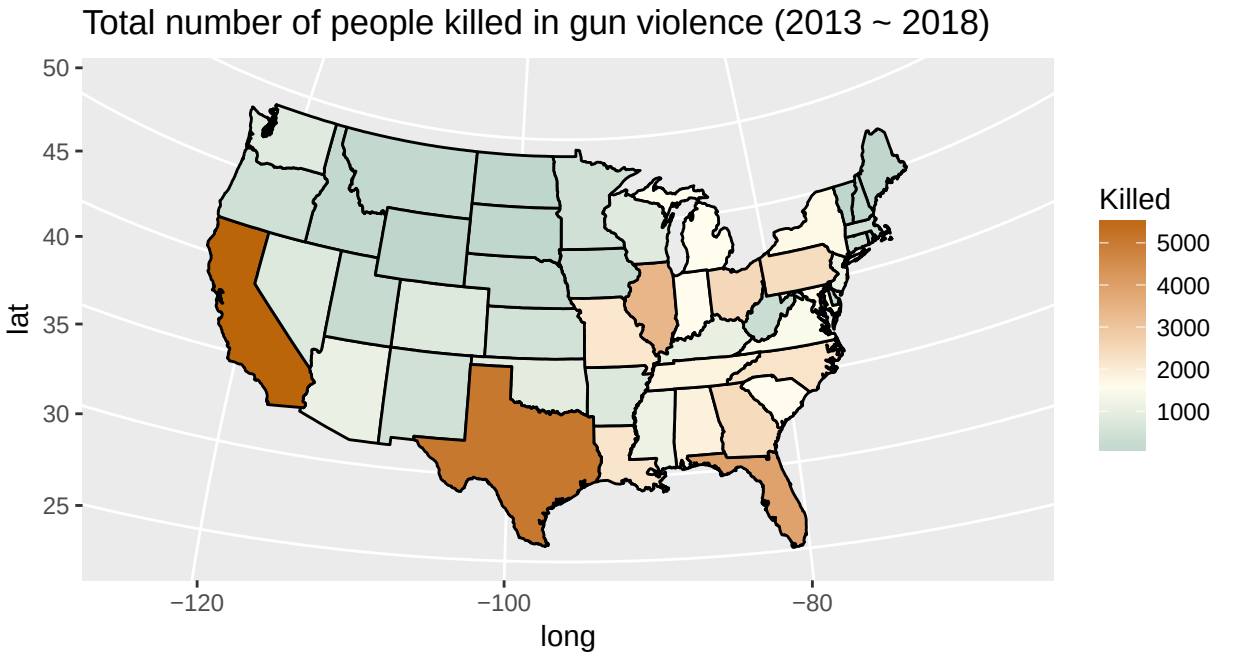
```
## # A tibble: 51 x 3
##    state                Killed Injured
##    <chr>                 <int>   <int>
##  1 alabama                1880    2998
##  2 alaska                  267     325
##  3 arizona                1094    1096
##  4 arkansas                773    1347
##  5 california             5562    7644
##  6 colorado                796    1133
##  7 connecticut             341    1258
##  8 delaware                217     853
##  9 district of columbia    459    1415
## 10 florida                3909    7072
## # ... with 41 more rows
```

```
sortTotKilled <- gv_states2[order(-gv_states2$Killed), ]
sortTotKilled <- cbind(sortTotKilled, TotKilledRank = seq(from=1, to=51))

sortTotInjured <- gv_states2[order(-gv_states2$Injured), ]
sortTotInjured <- cbind(sortTotInjured, TotInjuredRank = seq(from=1, to=51))
```

Geospatial map plots of gun violence in US (2013 ~ 2018)

This first mapping shows which states have the overall higher gun deaths. using a gradient color fill provides a heat map of gun violence intensity

```
library(ggplot2)

g <- ggplot ( gv_map, aes(x=long, y=lat, group=group, fill = Killed ))
g <- g + geom_polygon(colour = "black") + coord_map("polyconic")
g <- g + scale_fill_gradient2(low  = "#559999", mid ="#FFFDEE",
                              high = "#BB650B", midpoint = median(gv_map$Killed))
g <- g + ggtitle("Total number of people killed in gun violence (2013 ~ 2018)")


print(g)
```



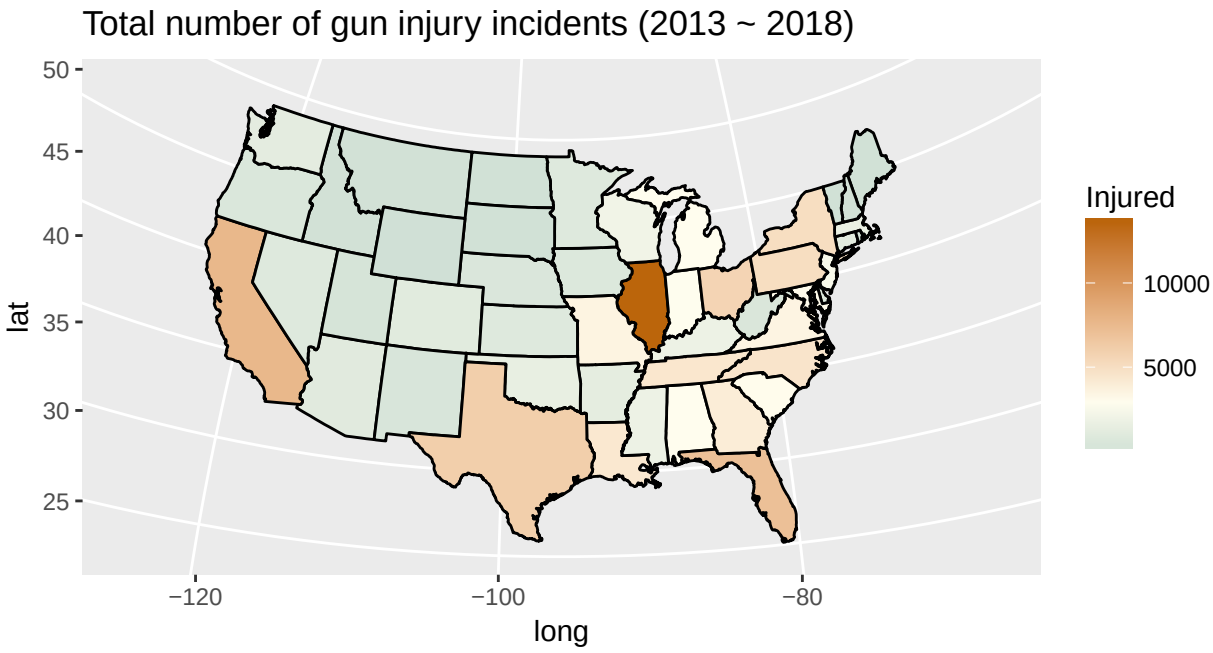
Top 5 states with highest total gun deaths

```
head(sortTotKilled,5)
```

```
##        state Killed Injured TotKilledRank
## 1 california   5562    7644             1
## 2      texas   5046    6106             2
## 3    florida   3909    7072             3
## 4   illinois   3409   13514             4
## 5       ohio   2508    5703             5
```

Second mapping shows then number of people injured in gun VIolence by state the same gratient color fill from white = low to dark amber = High

```
g <- ggplot ( gv_map, aes(x=long, y=lat, group=group, fill = Injured))
g <- g + geom_polygon(colour = "black") + coord_map("polyconic")
g <- g + scale_fill_gradient2(low  = "#559999", mid ="#FFFDEE",
                              high = "#BB650B", midpoint = median(gv_map$Injured))
g <- g + ggtitle("Total number of gun injury incidents (2013 ~ 2018)")

print(g)
```



Top 5 states with highest gun injuries

```
head(sortTotInjured,5)
```

```
##        state Killed Injured TotInjuredRank
## 1   illinois   3409   13514              1
## 2 california   5562    7644              2
## 3    florida   3909    7072              3
## 4      texas   5046    6106              4
## 5       ohio   2508    5703              5
```

Load census 2013 population data

By dividing the reported number of incidents per state with the 2013 US census data, the rate of gun violence per capitia is normalized in the data.

```
#
#census population data extracted manually from
#http://www.enchantedlearning.com/usa/states/population.shtml
# saved locally into a csv table

census <-read.csv("census.csv", header = TRUE, sep = ",")
census$State <- tolower(census$State)

head(census,5)
```

```
##        State Ranking Population
## 1 california       1   38332521
## 2      texas       2   26448193
## 3   new york       3   19651127
## 4    florida       4   19552860
## 5   illinois       5   12882135
```

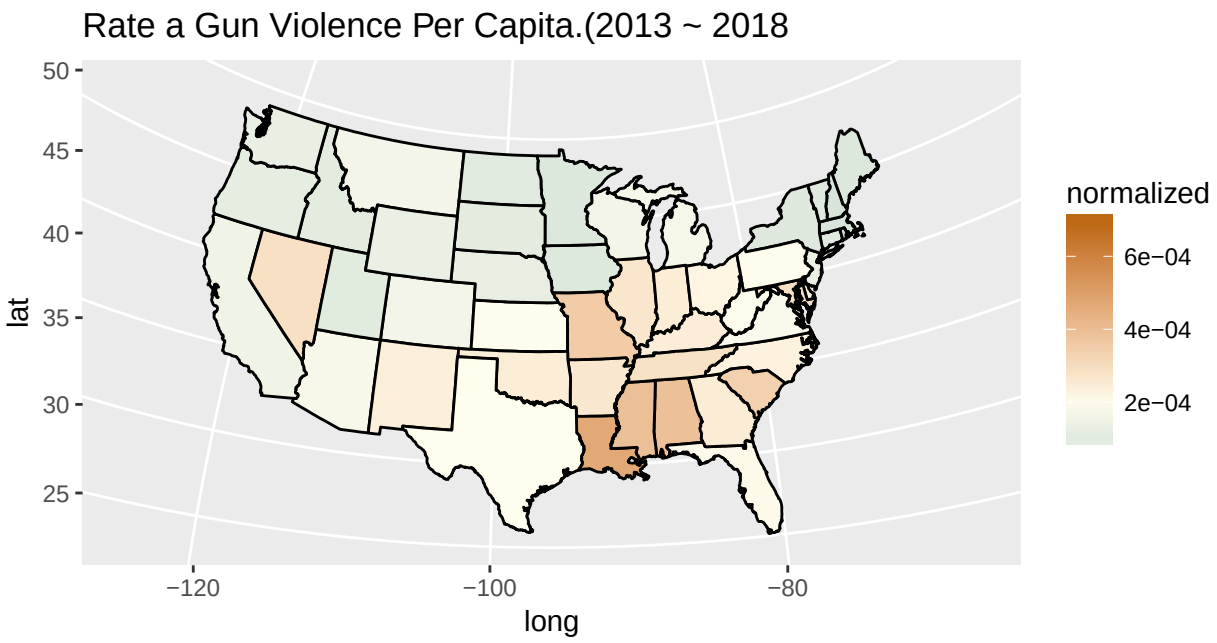
```
## just an intermediate working table to merge in census data columns into
gv_map_norm <- merge(gv_map, census,by.x = "region" ,by.y = "State" )

#create new dataset with normalized violence based on state population
gvmn2 <- cbind(gv_map_norm,(gv_map_norm$Killed/gv_map_norm$Population))
names(gvmn2) <- c("region", "long", "lat", "group", "order", "subregion",
                  "Killed", "Injured", "Ranking","Population", "normalized")



sortedgv <-  gvmn2 %>% group_by(region) %>% summarise(Killed = mean(Killed), Injured = mean(Injured), P
sortedgv <- sortedgv[order(-sortedgv$PerCapita), ]
```

Map plot of gun violence normalized by population of each state.

This map highlights the states with the highest per capitia gun violence. The same color gratient is applied to show the shift in gun violence intensity based on population

```
g <- ggplot(gvmn2, aes(x=long, y=lat, group=group, fill = normalized))
g <- g + geom_polygon(colour = "black") + coord_map("polyconic")
g <- g + scale_fill_gradient2(low  = "#559999", mid ="#FFFDEE",
                              high = "#BB650B", midpoint = median(gvmn2$normalized ))
g <- g + ggtitle("Rate a Gun Violence Per Capita.(2013 ~ 2018")
```

```
print(g)
```




```
sorted2 <- cbind(sortedgv, PerCapitaRank = seq(from=1, to=49))
```

Highest gun violence states when adjusted per capita

```
head(sorted2, 5)
```

```
##                region Killed Injured    PerCapita PerCapitaRank
## 1 district of columbia    459    1415 0.0007100328             1
## 2           louisiana   2179    4398 0.0004710873             2
## 3         mississippi   1176    1883 0.0003931523             3
## 4             alabama   1880    2998 0.0003889342             4
## 5            missouri   2136    3585 0.0003533983             5
```

Lowest gun violence states when adjusted per capita

```
tail(sorted2,5)
```

```
##           region Killed Injured    PerCapita PerCapitaRank
## 45     minnesota    461     916 8.504939e-05            45
## 46         maine    112     132 8.431817e-05            46
## 47 massachusetts    472    1701 7.052329e-05            47
## 48 new hampshire     88     144 6.649243e-05            48
## 49  rhode island     63     346 5.991378e-05            49
```